

Kathrin Steyer

USUELLE WORTVERBINDUNGEN DES DEUTSCHEN

Linguistisches Konzept und lexikografische Möglichkeiten

Abstract

Der Artikel schlägt ein für lexikografische Zwecke adaptierbares linguistisches Modell von üblichen Wortverbindungen vor, das die verschiedenen Herangehensweisen der Idiomatikforschung integriert, das streng korpusbasiert ist und die Kontexte von Wortverbindungen konsequent einbezieht. Das Modul 'Usuelle Wortverbindungen des Deutschen' ist ein zentrales Konzept des IDS-Projektes „Wissen über Wörter“, ein hypertextbasiertes, lexikalisch-lexikologisches Informationssystem, das in seinem Endausbau circa 300 000 Stichwörter enthalten wird. Korpusstatistische Kookkurrenzanalysen stellen hierbei ein wichtiges lexikografisches Arbeitsinstrument für die Rekonstruktion von Lesarten, von semantischen Merkmalen und Eigenschaften der Lemmata dar. Usuelle Wortverbindungen (Kollokationen, Phraseologismen und andere nicht-idiomatische Wendungen) werden in diesem elektronischen Nachschlagewerk darüber hinaus selbst zum Gegenstand lexikografischer Beschreibung, zum einen als Kookkurrenzangaben zu jedem Einwortlemma und zum anderen in einem eigenständigen Artikeltyp 'Mehrwortlemma'. Schließlich bietet diese kookkurrenzbezogene Herangehensweise eine fundierte empirische Basis für linguistische Untersuchungen.

The article presents a linguistic model of frequent word combinations which is adaptable for lexicographical purposes, which integrates the different approaches of idiom research, which is strictly corpus-based, and which includes contexts of word combinations in a consistent way. The module 'Frequent word combinations in German' is a central concept of the IDS project „Wissen über Wörter“ (Knowledge about Words), a hypertext-based, lexical-lexicological information system, which when completed will contain approx. 300,000 entries. Statistical co-occurrence analyses here represent an important lexicographical instrument for the reconstruction of variants, semantic features and the characteristics of headworks. Moreover, in this electronic reference work frequent word combinations (collocations, idioms and other non-idiomatic expressions) are themselves given lexicographical descriptions, on the one hand in the form of co-occurrence information about each single-word entry and on the other hand in their own type of article „multiple-word entry“. Finally this co-occurrence-based approach offers a sound empirical base for linguistic research.

0. Vorbemerkung

Das im Folgenden vorgestellte Konzept usueller Wortverbindungen bildet einen Schwerpunkt im Rahmen des IDS-Projektes „Wissen über Wörter“, ein lexikalisch-lexikologisches Informationssystem, das auf den mittlerweile ca. 650 Millionen Wörtern umfassenden IDS-Korpora basiert und den deutschen Wortschatz in umfassender Weise dokumentieren, allgemeinverständlich erklären und linguistisch erläutern wird. In seinem Endausbau wird dieses elektronische Nachschlagewerk ca. 300 000 Stichwörter enthalten und durch seine Hypertextstruktur sowohl Experten als auch sprachinteressierten Laien eine flexible, auf individuelle Bedürfnisse abgestimmte Informationsauswahl und -tiefe über Internet und CD-ROM und – je nach Bedarf – als Spezialnachschlagewerke in Buchform anbieten können.¹ Das Informationssystem unterscheidet sich in Aufbau und Nutzungsmöglichkeiten fundamental sowohl von herkömmlichen **Printwörterbü-**

¹ Weitere Informationen auf der Projekt-Homepage unter <http://www.ids-mannheim.de/wiw/> und in folgenden Publikationen: Fraas/Haß-Zumkehr 1998; Haß-Zumkehr 2000.

chern als auch von Wörterbüchern, die zwar in eine elektronische Form gebracht wurden, aber trotzdem als (lineares) Printmedium strukturiert sind.

1. Eine Bestandsaufnahme

1.1 Linguistische Ansätze

Ein Ergebnis korpusbasierter linguistischer Forschung der vergangenen Jahre ist die Erkenntnis, dass Sprachgebrauch stark routiniert ist und sich Textproduktion in ungleich höherem Maße über das Abrufen relativ stabiler Wortverbindungen realisiert als bisher angenommen. Darauf verweist Stubbs und schreibt der idiomatischen Kompetenz des Sprechers eine ebenso wichtige Rolle für ein kulturell angemessenes Kommunizieren zu wie dem Beherrschen sprachlicher Regularitäten, z.B. grammatischer Regeln (vgl. Stubbs 1997, S. 153). Bei einem hypertextuellen Nachschlagewerk, das den deutschen Wortschatz unter Berücksichtigung aller Ebenen des Sprachsystems und aller Aspekte des Sprachgebrauchs in umfassender Weise dokumentieren wird, rücken somit auch Fragen der Idiomatik ganz zwangsläufig ins Zentrum des Interesses. Schon in der ersten Konzeptionsphase des Projektes wurde deutlich, dass ein für lexikografische Zwecke adaptierbares Modell von üblichen Wortverbindungen entwickelt werden muss, das die verschiedenen Herangehensweisen der Idiomatikforschung integriert² und sie in ein neues Licht stellt. Betrachtet man die Forschungslandschaft auf diesem Gebiet, werden einige Widersprüche und Desiderate sehr deutlich: Es gibt eine Vielzahl disparater Ansätze und Terminologien, die zumeist ziemlich beziehungslos nebeneinander existieren oder sich sogar in Abgrenzung voneinander definieren.³ Die Gründe für diese Vielfalt liegen auf der Hand: Fast jede linguistische Richtung, von der generativen Transformationsgrammatik über die kontrastive Linguistik bis hin zur Stilistik, stößt irgendwann auf das Problem, dass es zwischen Wörtern neben rein grammatischen Verbindbarkeiten weitere strukturelle Beziehungen gibt, die es zu analysieren und zu beschreiben gilt. Jede Richtung hat dafür ein eigenes Modell entwickelt und zumeist genügt dies auch: Wortübergreifende Phänomene werden als Phraseologismen **oder** als Kollokationen **oder** als Valenz **oder** als syntaktische Kookkurrenz **oder** als Funktionsverbgefüge **oder** als Regelsystem semantischer Vereinbarkeiten **oder** als kommunikative Formeln betrachtet.⁴ Im Kontext eines komplexen Informationssystems ist es jedoch unabdingbar, die verschiedenen Perspektiven zusammenzuführen. In besonderer Weise betrifft dies die Verbindung der eher klassifikatorisch-analytischen Ansätze der Phraseologie mit den statistisch-syntagmatischen Methoden der Kollokationsanalyse, wie sie u.a. von der Schule des britischen Kontextualismus⁵ entwickelt wurden.

² Idiomatikforschung sei hier als Oberbegriff für all jene Richtungen verwendet, die sich – auf welche Weise auch immer – mit festen Wortverbindungen bzw. Regularitäten der präferierten Kombination von Wörtern beschäftigen.

³ Vgl. z.B. Hausmanns Abgrenzung von den statistischen Methoden des britischen Kontextualismus (Hausmann u.a. 1985, S.124ff.).

⁴ Ausnahmen bilden die einschlägigen Überblicksmonografien von Burger/Buhofer/Sialm 1982; Palm 1995; Fleischer 1997 und Burger 1998. Aber auch sie fokussieren vor allem die Phraseologie als Kern des Phänomens der festen Wortverbindungen. Zur terminologischen Vielfalt vgl. Donalies 1994.

⁵ Instruktive Überblicke über die Schule des britischen Kontextualismus und andere relevante Kollokationsrichtungen bieten Bahns 1996 und Lehr 1996.

Die **Phraseologie** gilt zu Recht als eine der elaboriertesten linguistischen Disziplinen überhaupt, mit einer schier unüberschaubaren Fülle von Arbeiten, einsprachigen bzw. kontrastiven Analysen und Klassifikationen.⁶ Bisher standen dabei vor allem jene festen Wortverbindungen im Mittelpunkt des Interesses, die u.a. als 'Idiome', 'Redewendungen', 'Phraseologismen' oder 'feste Fügungen' bezeichnet wurden (z.B. *das Kind mit dem Bade ausschütten*). Wenn in der Fachliteratur auch auf verschiedene Weise definiert, sind ihnen im Grunde folgende Eigenschaften gemeinsam: Solche Wortverbindungen weisen eine – zumeist aus einer sprachstrukturellen Anomalie resultierende – Festigkeit auf. Sie übernehmen die Funktion eines Einzelexems; ihre Gesamtbedeutung ist nicht direkt aus der Summe der Einzelbedeutungen ihrer Elemente ableitbar. Je geringer die Transparenz einer Wortverbindung bzw. je größer die Differenz zwischen wörtlicher und übertragener Bedeutung ist, desto idiomatischer ist sie im Verständnis dieser Richtung. In der neueren Literatur finden sich jedoch vermehrt Problematisierungen des tradierten, nun als zu starr angesehenen Phraseologiebegriffs. So konstatiert Burger, dass die phraseologische Forschung überzeugend klar gemacht habe, dass Phraseologismen polylexikalische Einheiten von **relativer** Festigkeit seien, die untereinander und mit dem Wortlexikon auf vielfältige Weise verknüpft sind (1998, S. 31).

„Statt dessen brauchen wir eine flexible Konzeption von Phraseologie, die den Gedanken toleriert, daß man innerhalb eines Phraseologismus stabilere und weniger stabile Elemente annehmen darf, ohne daß damit sein Status als linguistische bzw. psycholinguistische Einheit verloren ginge.“ (ebd.)

Ein weiteres Indiz für die Hinwendung zu einer dynamischeren Phraseologieauffassung ist die wachsende Zahl von Untersuchungen und Analysen, die den sprachlich-kommunikativen Kontext von Idiomen in die Analyse einbeziehen und die Festigkeit von Wortverbindungen nicht mehr nur an syntaktisch-semantische Restriktionen oder Defekte binden (vgl. dazu vor allem Sandig 1994, Stein 1995, Gréciano/Rothkegel 1997, Wirrer 1998).

Die **Kollokationsforschung** ist dagegen eine vergleichsweise junge Disziplin, die ihre Wurzeln in mehreren Forschungsrichtungen hat: in der statistisch-syntagmatisch vorgehenden Schule des britischen Kontextualismus (vgl. u.a. Sinclair 1991), in den eher philologisch orientierten, aus der kontrastiven Linguistik und Lexikografie erwachsenden Kollokationsmodellen (vgl. vor allem Hausmann u.a. 1984, 1985, 1989, 1993; auch Bahns 1996; Benson/Benson/Ilson⁷ 1997) und nicht zuletzt in den Forschungen zu semantischen Vereinbarkeitsbeziehungen und begrifflichen Relationen (vgl. vor allem Agricola u.a. 1982, 1992, Viehweger 1977, 1982, 1989).

Die Schule des britischen Kontextualismus, die – beginnend bei Firth – als Begründerin der Kollokationsforschung gilt, stellt die statistische Signifikanz in den Mittelpunkt. Ein wichtiges Kriterium für Kollokationen ist hier die statistische Erwartbarkeit des Mitein-

⁶ Auf die Gesamtheit der Forschungsliteratur kann an dieser Stelle nicht eingegangen werden.

⁷ Benson/Benson/Ilson (1997) unterscheiden zwischen 'lexical collocations' und 'grammatical collocations'. Ihr Ansatz weist eine große Nähe zur Dichotomie 'grammatische Verknüpfung' vs. 'Bedeutungsverknüpfung' bei Agricola (1992) auf, ohne dass sich jedoch diese Ansätze bereits explizit aufeinander beziehen. Folgende Typen von 'lexical collocations' schlagen Benson/Benson/Ilson vor: Verb-Nomen/Pronomen, Verb-Nomen, Adjektiv-Nomen, Nomen-Verb, Nomen-Nomen, Adverb-Adjektiv, Verb-Adjektiv. Diese Typen decken sich wiederum in fast allen Punkten mit den Kollokationstypen von Hausmann (1989).

andervorkommens von Wortverbindungen im Korpus. Sinclair versteht unter 'collocation' „the occurrence of two or more words within a short space of each other in a text.“ (1991, S. 170). Er geht unter Verwendung statistischer Methoden – basierend auf großen Korpora – streng syntagmatisch vor, wobei für ihn 'Kollokation' den Oberbegriff für alle sich wiederholenden Kookkurrenzen (Idiome, Klischees, Phraseologismen, Komposita usw.) darstellt. Durch diese sehr weitgehende Definition wird das entscheidende Problem der Statuszuordnung, also die Frage, welche Wortkombinationen Kollokationen sind und welche Differenzierungen sie aufweisen, u.E. nicht beantwortet. Das Verdienst des britischen Kontextualismus besteht aber unbestritten darin, nach tatsächlichen Kollokationsvorkommen in „realen Texten“ zu fragen und Methoden entwickelt zu haben, mit deren Hilfe über große Textmengen recherchiert werden kann.

Neben der angelsächsischen Tradition ist es vor allem Hausmann, der die Relevanz von Kollokationen als „Halbfertigprodukte der Sprache“ (Hausmann 1985, S. 118) erkennt und sie ins Bewusstsein der Forschungsgemeinde rückt. Kollokationen werden zunehmend nicht mehr nur als unscheinbare kleine Schwestern der Phraseologismen betrachtet, denen man zwar irgendwie ständig begegnet, mit denen man aber nicht so richtig etwas anzufangen weiß. Hausmann unterscheidet in seinem Modell (u.a. 1984) zwischen fixierten Wortverbindungen wie Komposita und feste Redewendungen (*chambre forte*, *laver la tête*) und nicht-fixierten Wortverbindungen, die sich in Konter-Kreationen (*jour fissuré*), Kollokationen (*mur fissuré*) und Ko-Kreationen (*regarder un arbre*) unterteilen. Kollokationen (*schütteres Haar*) versteht er als typische, spezifische und charakteristische Zweierkombinationen, die aus einer Basis (*Haar*) und einem Kollokator (*schütter*) bestehen. (1985, S. 118; vgl. dazu ausführlich 2.) Die Kenntnis der Kollokationen sei nicht etwa sprachlicher Luxus, so Hausmann, sondern elementar, da in den Kollokationen in besonderer Weise die unverwechselbare Eigenart einer Sprache niedergelegt sei (1993, S. 9). Dies lässt sich am Beispiel bilingualer Kommunikation belegen: Phraseologismen wurden – im Gegensatz zu Kollokationen – bisher oft als Problemfall mehrsprachiger Kommunikation thematisiert, da ein L2-Sprecher zumeist ein Nachschlagewerk benutzen muss, um sie verstehen zu können. Dies führte zu einer Überschätzung von phraseologischen Wendungen für mehrsprachiges Kommunizieren und zu einer Unterschätzung der Bedeutung von Kollokationen. Auch wenn Phraseologismen nicht unwesentlich die Vielfalt und Eigenheit einer Sprache bestimmen, kann ein Nichtmuttersprachler aber dennoch auf sie verzichten und sich trotzdem angemessen in der Fremdsprache ausdrücken. Er kann sie einfach ‚vermeiden‘. In der Regel wird eine entwickelte phraseologische Kompetenz von ihm gar nicht erwartet, da immer die Gefahr von Normverletzungen oder kommunikativen Missverständnissen besteht. Ganz anders stellt sich dies bei Kollokationen dar: Ohne sie kommt der L2-Sprecher bei der Textproduktion in keinem Fall aus. Rezipiert ein Sprecher einen fremdsprachigen Text, macht es ihm in der Regel keine Schwierigkeiten, eine Fügung wie *to do an attempt* (einen Versuch machen) zu verstehen. Will er aber die deutsche Wendung ins Englische übersetzen, wird er möglicherweise **to make an attempt* sagen. Die Kombination *make-attempt* ist zwar verständlich, aber sie ist im englischen Sprachgebrauch nicht üblich. An den Kollokationen wird nicht unwesentlich der Grad der Beherrschung der jeweiligen Fremdsprache deutlich, obwohl ein Verstoß gegen eine 'Kollokationsnorm' eigentlich keine derart schwerwiegenden Folgen für den Kommunikationsverlauf hat wie eine 'falsche' (weil wörtlich übersetzte oder falsch angewendete) Redewendung. Programmatisch erscheint

demzufolge der Titel eines Artikels von Hausmann, in dem er sein Kollokationsmodell anhand der Analyse einer Übersetzung aus dem Französischen ins Deutsche entwickelt: „Wortschatzlernen ist Kollokationslernen.“ (1984, S. 395).

Dezidiert als Hausmann schlägt Viehweger, aufbauend auf Agricolas Konzept der „Wörter und Wendungen“ (vgl. u.a. 1992), dem Modell mentaler Wissensorganisation von Klix (u.a. 1984) und seinen eigenen Forschungen zur semantischen Analyse (vgl. u.a. 1977) eine Kollokationsdefinition vor, die versucht, die strukturellen Beziehungen zwischen Einheiten im Lexikon zu erhellen. Er versteht Kollokationen als zweigliedrige semantische Relationen zwischen Autosemantika, die in Abgrenzung von Phraseologismen regel- und normgerechte Verknüpfungen von Wortschatzelementen darstellen (1989, S. 889). Diese Definition erscheint uns in sprachtheoretischer Hinsicht am konsequentesten zu sein. In Hinblick auf die lexikografische Repräsentation plädiert er dafür, diese semantischen Vereinbarkeiten – im Gegensatz zu den bislang präferierten paradigmatischen Angaben – nicht länger als zufälliges Accessoire eines Wörterbucheintrages zu betrachten, da sie einen Datentyp darstellen, der einem vorrangigen Benutzerbedürfnis entspricht (ebd., S. 892).

Gegenwärtig rückt die Kollokationsthematik mehr und mehr auch als generelles sprachtheoretisches Problem in den Fokus der linguistischen Forschung. Einen wesentlichen Impuls liefert hierbei Feilkes Konzept der ‚idiomatischen Prägung der Sprachkompetenz‘ (u.a. 1996).

1.2 Lexikografische Desiderate

Betrachtet man die Wörterbuchlandschaft im deutschsprachigen Raum in Hinblick auf die Aufbereitung von Mehrwortangaben, so muss man zunächst konstatieren, dass diese zwar in jedem Wörterbuch vorkommen, aber zumeist völlig unsystematisch. So wird z.B. in der Regel nicht zwischen Kollokationsangaben und phraseologischen Angaben unterschieden. Burger begründet diese fehlende Systematik mit der stiefmütterlichen Behandlung der Phraseologie in den gängigen Lexika der deutschen Sprache und vor allem mit dem fehlenden theoretischen Konzept für eine moderne Phraseografie (vgl. 1998, S.168f.), was sich ohne Probleme für die Mehrwortlexikografie generell verallgemeinern lässt.⁸ Trotz der Vielfalt der Ansätze im Bereich der Idiomatikforschung existiert bisher also kein für lexikografische Zwecke adaptierbares Modell von gebräuchlichen Wortverbindungen einer Sprache, das auf einer konsequenten Analyse von tatsächlichem Sprachgebrauch beruht und die **kontextuellen** Bedingungen systematisch einbezieht.

Folgende Gruppen lassen sich innerhalb der Idiomatiklexikografie unterscheiden:

- Mehrwortangaben in einsprachigen Wörterbüchern,
- Mehrwortangaben in zweisprachigen Wörterbüchern,
- idiomatische einsprachige und kontrastive Spezialwörterbücher.

Bei den Spezialwörterbüchern nehmen die phraseologischen Sammlungen den eindeutig größten Platz ein. Dass diese hochidiomatischen, wenig transparenten Wendungen in un-

⁸ Zu Recht sieht er aber einen Fortschritt mit dem Erscheinen der idiomatischen Spezialwörterbücher von Schemann (1993) und vor allem Duden 11 (1998) (ebd.).

gleich höherem Maße Beachtung fanden, sowohl in der wissenschaftlichen Analyse als auch bei der lexikografischen Erfassung und Bearbeitung, liegt in der sprachlichen Natur von Phraseologismen begründet: Phraseologismen bilden zumeist relativ eindeutige Fälle von Idiomatizität ab und haben einen relativ überschaubaren Bestand an sprachlichen Formen, die sich syntaktisch und semantisch – auch ohne Korpora und Computerverfahren – gut klassifizieren lassen. Sie bieten auf Grund ihrer mangelnden Transparenz genug Stoff für lexikologische Analysen und sind dazu prädestiniert, in ein Wörterbuch aufgenommen zu werden. Sieht man von Agricola 1992, Schemann 1993 und dem Duden 11 ab, finden Kollokationen des Deutschen bisher fast ausschließlich in allgemeinsprachigen Wörterbüchern Eingang, jedoch ohne ein theoretisch fundiertes Kollokationsmodell. Dass diese lexikografischen Bearbeitungen mit einer Vielzahl von offenen Problemen verbunden sind, verdeutlicht Burger, dessen kritische Bilanz bisheriger lexikografischer Bearbeitung von Phraseologismen als symptomatisch für den gesamten Bereich der Mehrwortangaben gelten kann (1998, S. 170ff.). An dieser Stelle seien vier dieser Problemkreise näher betrachtet, die uns elementar erscheinen und immer auch generelle Probleme der lexikografischen Arbeit betreffen.

Das erste Problem, vor dem Lexikografen stehen, ist das Problem der **Statuszuordnung**, also der generellen Entscheidung, welcher Phrase überhaupt der Status einer festen Wortverbindung zugesprochen werden kann und unter welchen Bedingungen sie ein Lemmakandidat wird. Diese Zuordnungen werden bisher zumeist intuitiv, auf der Basis individueller Belegsammlungen vorgenommen und sind nicht konsequent korpusbasiert. Auch wenn die klassischen 'Zettelkästen' zunehmend durch computergestützte Belegsammlungen ersetzt werden, entscheidet der Lexikograf in den meisten Fällen nach wie vor auf Grund seiner Sprachkompetenz, ob er eine Variante für einschlägig und relevant hält und – in unserem Fall – ob eine Wortverbindung typischer als eine andere ist. Für die neueren Ausgaben von Wörterbüchern werden schon Korpora genutzt, aber nur – das räumen die Lexikografen zumeist selbst explizit ein (z.B. Zimmer 1990, Schemann 1993) – um die eigenen, zwar auf jahrelanger Wörterbucherfahrung basierenden, dennoch aber subjektiven Entscheidungen für einen Eintrag zu verifizieren bzw. um die Korpusbelege zur Illustration der Artikel zu nutzen. Das führt dann dazu, dass in die Wörterbücher und Sammlungen zumeist die auffälligen, geradezu klassischen Beispiele und Belege aufgenommen werden. Es gibt mittlerweile so etwas wie ein kollektives Beispielgedächtnis von Forschern und Lexikografen, dessen Wurzeln sich teilweise über Jahrzehnte zurückverfolgen lassen und das einer empirischen Überprüfung am aktuellen Sprachgebrauch bedarf. Der Ausweg kann nur eine strikt korpusbasiert-empirische Fundierung der Lexikografie sein. Konsequente Korpusbasiertheit bedeutet aber immer noch eine 'psychologische' Barriere: Der Lexikograf sieht sich bei seinen Recherchen nicht selten mit Befunden konfrontiert, die seiner eigenen Sprachintuition widersprechen, aber trotzdem als repräsentativ für den Sprachgebrauch angesehen werden müssen.

Das zweite Problem ist ebenso gravierend: Es muss entschieden werden, unter welchem **Stichwort** die festen Wortverbindungen in den Printwörterbüchern aufbereitet werden. Zumeist bestehen diese Phrasen ja aus mehr als einer sinntragenden Einheit. Dies wird – wie Burger meint – zumeist 'pragmatisch' gelöst, indem idiomatische Angaben unter der ersten sinntragenden Einheit oder auch unter dem beteiligten Substantiv abgehandelt werden, bei den anderen beteiligten Basiselementen würde man nur einen entsprechenden

Verweis finden. Mit seiner Einschätzung, diese Praxis für ein „relativ unproblematisches Verfahren“ zu halten, stimmen wir allerdings nicht überein (ebd., S.171f.). Bei Kollokationsangaben müsste dies folgerichtig unter der Basis erfolgen, die oft durch ein Nomen repräsentiert ist. Wie fragwürdig diese Einordnungen sind, macht ein Beispiel deutlich, das wir von Hausmann aufgreifen: Die Wortkombination *schütteres Haar* würde man als Kollokationsangabe unter der 'Basis' *Haar* finden, obwohl das Nachschlagebedürfnis zum 'Kollokator', dem Lexem *schütter*, wohl viel größer sein dürfte. In diesem einfachen Fall *schütteres Haar* lässt sich das Problem wohl noch mit einer doppelten Behandlung sowohl unter *Haar* als auch unter *schütter* bzw. mit entsprechenden Querverweisen lösen. Bei komplexeren Phrasen ist dies auf Grund des begrenzten Umfangs der Wörterbuchartikel sicher nicht möglich. Hausmann selbst hebt die Praxis z.B. des DUW hervor, Kollokationen in vielen Fällen sowohl unter der Basis als auch unter dem Kollokator zu verzeichnen, räumt aber ein, dass dies aus platzökonomischen Gründen nicht durchweg möglich ist (1985, S. 121ff.). Auch hier findet sich ein Problem bisheriger, d.h. gedruckter Wörterbücher. Der klassische Konflikt beim lexikografischen Arbeiten zwischen dem Zwang zur Komprimierung einerseits und dem Anspruch möglichst umfassender Information andererseits führt zu permanenten Entscheidungen für oder gegen eine lexikografische Angabe. Die vorhandenen Wörterbücher können deshalb immer nur eine Auswahl solcher typischen Kombinationen bieten und diese Belege nur in einer einmal festgelegten Ordnung präsentieren. Dies ist u.a. immer dann problematisch, wenn es um die Entscheidung geht, welche Angaben genereller Natur sind und welche zu den einzelnen Lesarten eines Lemmas gehören.

Die Folgen des Komprimierungszwangs in der papiergebundenen Lexikografie zeigen sich vor allem auf dem Gebiet der **pragma-stilistischen Angaben**, die gerade für idiomatische Angaben auf Grund des hohen Grades an Kontextgebundenheit von besonderer Relevanz sind. Die vorherrschenden Etikettierungen und Zuordnungen zu Stilebenen (vgl. z.B. 'gehoben' vs. 'umgangssprachlich') reduzieren sich auf wenige Wörter. Zum Teil sind es Einwortkommentare, die keinesfalls dem modernen Forschungsstand der Pragmatik, Textlinguistik und Stilistik entsprechen. Diese Einwortkommentare suggerieren, dass es feste, den Lexemen inhärente pragmatische Merkmale und Konnotationen gäbe. Dabei sind gerade diese Angaben in hohen Maße kontextabhängig und werden einem Lexem variabel zugeschrieben, was sich aber nur in satzwertigen Kommentaren angemessen darstellen lässt. Die Einwortmarker werden in der Metalexikografie deshalb zu Recht zunehmend kritisiert, ohne dass sich dies bisher in der praktischen Wörterbucharbeit niedergeschlagen hätte.

Ein weiteres, u.E. bisher absolut ungelöstes Problem, das Burger diskutiert, ist die Frage nach der **Normalform** eines Phraseologismus. In gewisser Weise betrifft dies auch Kollokationsangaben, wie zu zeigen sein wird. Die Normalformen, also die auf den obligatorischen (nicht-variablen) Teil einer Wortverbindung reduzierten Grundformen, stellen Konstrukte für lexikografische Zwecke dar und differieren in den bisherigen Wörterbüchern und idiomatischen Sammlungen folgerichtig beträchtlich. Auch hier kann letztendlich nur eine Korpuserhebung Aufschluss darüber bringen, wie die Grundform beschaffen ist und welche Modifikationen tatsächlich möglich sind.

Viele Desiderate der klassischen (Print-)Lexikografie können durch die modernen Möglichkeiten der Computlexikografie beseitigt werden. So gibt es vermehrt Versuche, vor

allem Kollokationen, aber auch Phraseologismen aus Korpora maschinell aufzubereiten und abrufbar zu machen. Gerade bei denjenigen Forschungsvorhaben, die sich zum Ziel gesetzt haben, das Kollokationsaufkommen von Sprachen mit statistischen Methoden zu erheben und maschinell aufzubereiten, ist jedoch bisher stets eine starke Orientierung auf die Entwicklung entsprechender **automatischer Tools** zu verzeichnen.

2. Das Konzept 'Usuelle Wortverbindungen'

Das von uns vorzuschlagende Konzept der 'Usuellen Wortverbindungen' kann in Verbindung mit den neuen Möglichkeiten hypertextbasierter Informationssysteme und dem Prinzip der strengen Korpusbasiertheit viele der eben beschriebenen linguistischen und lexikografischen Probleme lösen.

Wir unterscheiden in diesem Modell zunächst grundsätzlich zwischen okkasionellen und usuellen Kookkurrenzen (usuellen Wortverbindungen). **Okkasionelle Kookkurrenzen** stellen temporäre Wortverbindungen dar, die sich erst im Sprachproduktionsprozess konstituieren und denen kein über die aktuelle Situation hinausgehender Einheitsstatus zugeschrieben werden kann. **Usuelle Wortverbindungen** – und nur die sind für unser Informationssystem einschlägig – sind über das Einzelwort hinausgehende sprachliche Erscheinungen, die als komplexere Einheiten reproduziert werden können und deren Elemente einen höheren Wahrscheinlichkeitsgrad des Miteinandervorkommens besitzen, als das bei okkasionellen Wortverbindungen der Fall ist. Diese Charakterisierung sagt zunächst weder etwas über die sprachebenenabhängige Determiniertheit (also primär syntaktischer, semantischer oder pragmatischer Natur) noch über eine möglicherweise vorhandene idiomatische Komponente aus. **Usueller Status** ist für uns **nicht** – und hier unterscheiden wir uns von den meisten bisherigen Ansätzen – an das Kriterium der **Idiomatizität** gebunden.⁹ **Usuell** sind für uns alle jene Wortkombinationen, die eine historisch gewachsene Gebrauchsnorm repräsentieren, also Standardverwendungen darstellen und in diesem Sinne **typisch** sind (vgl. dazu Schmidt u.a. 1991, 1995). Nicht die Abweichung, sondern die Norm ist für uns also das entscheidende Kriterium.

Wortverbindungen sind in unserem Verständnis dann typisch, wenn sie auf Grund des usuellen Gebrauchs einen inneren Bindungsgrad der Komponenten aufweisen, der diese nicht beliebig austauschbar macht. **Usueller Gebrauch** ergibt sich für uns aus der statistischen Signifikanz der Wortverbindung im Korpus.¹⁰ Mit der Bestimmung des usuellen Status im Sinne einer **Gebrauchsnorm** ist es möglich, sich aus dem Korsett der tradierten Grenzziehungen zu befreien. Auf diese Weise können sprachliche Erscheinungen wie Kollokationen einerseits und Phraseologismen andererseits in ein einheitliches Modell integriert werden.¹¹ Das Konzept der usuellen Wortverbindungen ermöglicht auch

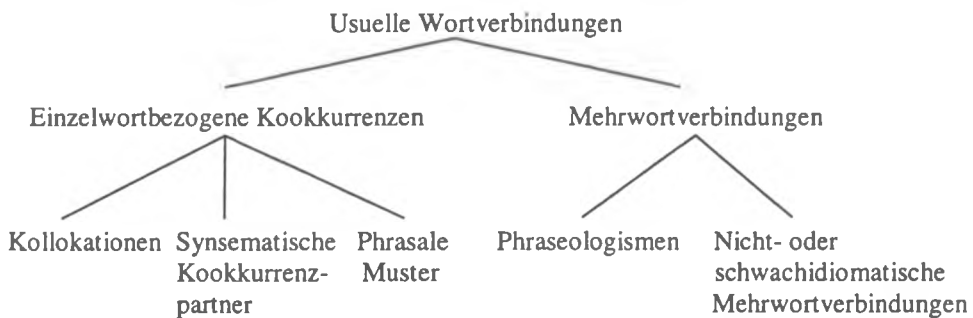
⁹ Auch Hausmanns Beispiele für Kollokationen (in Abgrenzung zu freien Ko-Kreationen) weisen immer eine gewisse idiomatische Komponente auf, obwohl er dies nicht als explizites Kriterium ansetzt.

¹⁰ Statistische Signifikanz bedeutet der Grad der lexikalischen Kohäsion zwischen zwei Lexemen.

¹¹ In der Phraseologie ist eine gewisse Unentschlossenheit bei der Einordnung der Kollokationen zu verzeichnen. Zumeist werden sie trotz Einschränkungen dann doch den 'Phraseologismen' zugeordnet. Fleischer versteht sie vor allem als Nominationsstereotype, wobei er in der 2. Auflage der 'Phraseologie' auch qualitative Unterschiede zwischen beiden Konzepten heraushebt: Während er Kollokation als strukturell restringierte Verbindungen der Ebene der langue zurechnet, lassen sich seiner Auffassung nach Nominationsstereotype nicht darauf reduzieren, da sie ihren stereotypen Charakter oft auch als Er-

die Einbeziehung jener typischen Wortkombinationen, die an keinerlei Idiomatik oder regelgeleitete semantische Verbindbarkeit gebunden sind, dennoch aber unverzichtbare Bausteine der Sprachproduktion und -rezeption darstellen.

Wir unterscheiden weiter zwischen einzelwortbezogenen usuellen Kookkurrenzen einerseits und usuellen Kookkurrenzen, die den Status von Mehrwortverbindungen haben, andererseits:



Einzelwortbezogene usuelle Kookkurrenzen umfassen Kollokationen, usuelle synsemantische Kookkurrenzrelationen und usuelle phrasale Muster in der Umgebung eines Wortes. Usuelle synsemantische Kookkurrenzen stellen ein primär grammatisches Phänomen dar, das an dieser Stelle nicht weiter behandelt wird. Auf die Relevanz useller phrasaler Muster kommen wir im 3. Punkt zurück.

Kollokationen definieren wir als binäre, usuelle Wortverbindungen, die aus zwei Auto- semantika bestehen. Sie stellen typische Verknüpfungen von Wortschatzelementen dar, die semantischer Natur sind, aber in variierenden syntaktischen Strukturen realisiert werden. Es geht also nicht nur um Kombinationen wie Hund-bellen, Hund-beißen oder Hund-bunt bzw. um Hund-Herrchen und Hund-Katze, sondern auch um usuelle Kookkurrenzrelationen wie Hund-Leinenzwang oder Hund-Postbote.¹² Dabei haben die aus den Korpora extrahierten Kollokationen qualitativ sehr unterschiedliche innere Bindungsqualitäten. Es ist daher zu unterscheiden zwischen jenen Zweiwortkombinationen, die neben der statistischen Signifikanz auch typische **sprachstrukturell motivierte** syntagmatische bzw. paradigmatische Vereinbarkeiten aufweisen wie Hund-bellen, Hund-bissig, Hund-Katze oder Hund-Pudel, und jenen statistisch signifikanten Kollokationen, die typisch im Sinne einer **korpusemantischen** bzw. **diskurspragmatischen** Relation sind, also signifikante diskursabhängige Kookkurrenzen wie Hund-Leinenzwang, Hund-Postbote und

scheinung der parole durch häufigen Gebrauch erhalten (1997, S. 252). Andere ordnen Kollokationen – basierend auf Fleischers Zentrum-Peripherie-Modell – als 'Periphere Phraseologische Einheiten' ein (vgl. u.a. Wojak 1994, S. 651ff.; Irsula Pena 1994, S. 27ff.). Inzwischen herrscht in der Phraseologie mehr und mehr Konsens darüber, den Terminus 'Kollokation' für all jene Phraseologismen zu verwenden, die nicht oder nur schwach idiomatisch sind.

¹² Wir verwenden im Folgenden mit dem Kollokationsfeld von *Hund* (u.a. mit der Relation Hund-bellen) bewusst einen 'Beispielklassiker' der Idiomatik (vgl. bereits Porzig 1934), weil sich auf diese Weise die neuen Möglichkeiten unseres Konzepts besonders nachvollziehbar verdeutlichen lassen.

Hund-Tierheim usw. Entscheidend ist in jedem Fall, dass die Kollokation als strukturelle Kombination angesehen wird, bei der die synsemantischen Komponenten für das Wesen dieser Relation zunächst nicht von Belang sind: Der Wortverbindung *Sie könnte immerzu in der Sonne sitzen* liegt die Kollokation Sonne-sitzen zu Grunde¹³, der Phraseologismus *wie Hund und Katze* beinhaltet die Kollokation Hund-Katze. Eine Zweiwortkombination ist also nicht deshalb eine Kollokation, weil sie eine idiomatische Komponente hat oder neben einer wörtlichen auch noch eine übertragene Bedeutung besitzt, sondern weil sie eine typische Verbindung repräsentiert. Sie ist in erster Instanz eben kein phraseologisches Phänomen, sondern ein usuelles. Dies bedeutet nicht, dass es nicht Kollokationen gibt, die eine idiomatische Komponente aufweisen. Das entscheidende Kriterium ist für uns aber die statistische Signifikanz einer Wortverbindung.¹⁴ So erscheinen auch die Klassifikationen von Hausmann in einem neuen Licht: Wenn er die Auffassung vertritt, dass *schütteres Haar* und *eingefleischter Junggeselle* Kollokationen seien, *regarder un arbre* dagegen eine Ko-Kreation, also eine freie Verbindung, möchte man dieser Einteilung aus reiner Sprachintuition sofort zustimmen. Was ist aber mit der Verbindung Baum-stehen? Im Sinne von Hausmann müsste man auch dieser Kombination einen freien, also nicht-kollokativen, Status zusprechen. Es ist nichts ‚Auffälliges‘ oder Spezifisches in dem Sachverhalt zu sehen, dass ein ‚Baum steht‘. Das Verb *stehen* ist aber ein statistisch signifikanter Kollokationspartner des Nomens *Baum* ebenso wie *pflanzen*, *fällen*, *wachsen*, *klettern*, *umstürzen*, *anketten*. Es handelt sich hier also um eine Gebrauchsnorm. Typischerweise wird dem Nomen *Baum* durch die Sprecher einer Sprachgemeinschaft u.a. die Eigenschaft ‘stehen’ zugeschrieben. Kommen wir noch einmal auf die Kombinationen *schütteres Haar* und *eingefleischter Junggeselle* zurück: *Schütteres Haar* besitzt durch die Komponente *schütter* ohne Zweifel eine größere Auffälligkeit. Einen Kollokationsstatus erlangt sie aber erst durch die statistische Erhebung. Das Nomen *Haar* ist tatsächlich der signifikanteste Kollokationspartner des Adjektivs *schütter*. Die Umkehrprobe funktioniert indes nicht: *Schütter* findet sich als Kollokationspartner von *Haar* auf dieser Skala im unteren Bereich; andere Adjektive wie *lang*, *kurz*, *blond*, *dunkel*, *schwarz* rangieren als typische Kollokationspartner von *Haar* dagegen ganz oben. Auch die Wortkombination *eingefleischter Junggeselle* scheint in unserem Sprachgedächtnis die prototypische Verbindung zu sein, was die entsprechenden Artikel zum Lemma *eingefleischt* im DUW, PAUL und WAHRIG zunächst bestätigen. Interessanterweise gibt SCHEMANN die Kollokation nur unter *Junggeselle an*¹⁵, DUDEN 11 und AGRICOLA nehmen sie überhaupt nicht auf:

¹³ Das Beispiel *in der Sonne sitzen* greift Burger von Feilke auf, vor allem als Beleg für den nicht an eine phraseologische Bedeutung gebundenen Usus einer Wortverbindung (1998, S.50f.). Mit seiner nicht weiter reflektierten Nebenbemerkung, dass es im Schweizerischen und Österreichischen durchaus üblich ist zu sagen, *an der Sonne ...*, gelangt Burger genau zu unserem Problem: Für die Bestimmung des usualen Gebrauchs in unserem Sinne ist die Verwendung verschiedener Präpositionen völlig unerheblich, in jedem Fall handelt es sich um die Kollokation Sonne-sitzen.

¹⁴ Für die statistische Analyse werden die am IDS von Cyril Belica entwickelten Kookkurrenzanalysetools angewendet (zur Funktionsweise vgl. auch Steyer/Teubert 1997). Alle folgenden statistischen Resultate basieren auf der Auswertung der IDS-Korpora.

¹⁵ Beim Lemma *Junggeselle* findet sich allein im DUW die entsprechende Kollokationsangabe *eingefleischter J.*

eingefleischt

DUW: [...] 1. die der angesprochenen Lebensweise, Eigenschaft o.ä. entsprechende *innere Einstellung ganz u. gar, durch u. durch verkörpernd* [u. in der Hinsicht *unverbesserlich*]: ein -er Junggeselle, Optimist. 2. zu nicht mehr änderbarer Gewohnheit, zur zweiten Natur geworden: ein -es Vorurteil; -e Sparsamkeit.

PAUL: [...] Es hat dann weiter die Bed. >unveränderlich, unverbesserlich< angenommen in *eingefleischter Junggeselle* [...]

WAHRIG: [...] in *Fleisch u. Blut eingewurzelt*; -e Gewohnheit; <fast nur in der Wendung> -er Junggeselle *J. aus Überzeugung, aus eigenem Willen*

Im Korpus gibt es im Gegensatz zu diesen sehr eindeutigen Aussagen auch bei der Kombination *eingefleischt-Junggeselle* nur eine partielle statistische Validierung. Statistisch signifikantester Kollokationspartner für *eingefleischt* ist das Nomen *Fan*. Das Nomen *Junggeselle* ist zwar auch statistisch signifikant belegt, aber nicht an erster Stelle. Andere Nomina, die in der Umgebung von *eingefleischt* belegt sind und noch vor *Junggeselle* rangieren, sind u.a.: *Anhänger, Cineast, Vegetarier, Gegner, Antikommunist*. Stellt das Nomen *Junggeselle* das Bezugsobjekt für eine Kollokationsangabe dar, haben *begeehrt, ewig, Ehebett, Braut, überzeugt, alternd* höhere Signifikanzwerte als *eingefleischt*. Bei einer Angabe im Wörterbuch müsste man zumindest die Skala der lexikalischen Kohäsion berücksichtigen und nicht nur eine Variante als die typische angeben. Hier zeigt sich u.E. das entscheidende Problem der Mehrzahl der bisherigen Klassifikationsmodelle der Idiomatikforschung: Sie haben alle mehr oder weniger versucht, solche Statuszuschreibungen per se vorzunehmen. Eine Wortverbindung ist auf Grund spezieller Eigenschaften (zumeist in ihrer Binnenstruktur) **an sich** eine Kollokation, ein Phraseologismus usw. Betrachtet man das Beispiel *eingefleischt-Junggeselle*, kann man ganz verschiedene Zuordnungen vornehmen: Man kann diese Wortverbindung auf Grund der idiomatischen Komponente natürlich als *Phraseolexem* einstufen und sie möglicherweise als zum Inventar deutscher Phraseologismen zugehörig definieren. Man kann ihr auch einen Kollokationsstatus im Sinne einer **aktuellen Gebrauchsnorm** zusprechen, da es für beide Partner eine statistische Belegung gibt, wenn man zugleich aber auch die gebräuchlicheren anderen Kollokationspartner aufführt und das Adjektiv *eingefleischt* nicht allein über die Kollokationsbeziehung mit dem Nomen *Junggeselle* definiert.¹⁶

Neben dem Überprüfen 'erwartbarer' Wortkombinationen besteht ein weiterer Vorteil der statistischen Methode darin, dass Kollokationen erfasst werden, die sich der Sprachintuition in viel größerem Maße verschließen, beispielsweise das typische Miteinandervorkommen zweier Nomina bzw. Verben innerhalb einer unmittelbaren sprachlichen Umgebung. Gerade diese Wortkombinationen, deren Elemente nicht unmittelbar syntaktisch aneinander gebunden sind, aber in einem engen sprachlichen Kontext häufig miteinander vorkommen, können in besonderer Weise die innere Struktur des Wort-

¹⁶ Auch eine angenommene unikale Komponente, also die Gebundenheit eines Elements an die jeweilige Wortverbindung bzw. das Nicht-Mehr-Vorkommen außerhalb des Phraseologismus lässt sich durch Korpuserhebung überprüfen. So meint Hausmann, dass *schütter* nicht ohne Bezug auf *Haar* und *eingefleischt* nicht ohne Bezug auf *Junggeselle* zu definieren sei (1985, S. 121). Die Korpuserhebung ergibt jedoch selbst bei *schütter* auch noch andere Referenzobjekte: *schütter besetzter Raum, schütter gewordener Schnee, schütterer Finanzen, schütterer Baumbepflanzung*.

schatzes erhellen. Darüber hinaus findet sich bei diesen nicht unmittelbar vor und nach einem Lexem vorkommenden Umgebungspartnern oft die Verbindung zu einer möglichen idiomatischen Wendung (vgl. dazu 3.).

Die Aussagekraft der statistischen Methode hängt natürlich im wesentlichen Maße von der **Qualität der zu Grunde liegenden Korpora** und der **Qualität der Abfrage-prozedur** ab. Wichtige Qualitätskriterien für Korpora sind u.a. die absolute Größe eines Korpus, die Breite des Textsortenspektrums und festlegbare Zeitfenster, die es ermöglichen, sprachliche Befunde auch zeitlich-historisch einordnen zu können.¹⁷ Darüber hinaus wird auch die statistische Methode durch linguistische Validierung ständig verfeinert und die Quote der Fehlbelege reduziert. Es wird schließlich empirisch zu überprüfen sein, wie viele Kollokationspartner, die für einschlägig gehalten werden, tatsächlich durch das Raster fallen und ob sie ohne statistische Belegung überhaupt eine Gebrauchsnorm darstellen, nur weil sie im Prinzip semantisch vereinbar sind und der Sprachintuition des Linguisten entsprechen.

Usuelle Mehrwortverbindungen besitzen im Gegensatz zu den einzelwortbezogenen Kookkurrenzen einen gewissen Einheitsstatus und können in einem weiten Sinne auch Einzelwortfunktionen übernehmen, da sie eine übersummativ Bedeutung bzw. kommunikative Funktion aufweisen. Den Kernbereich bilden hier Phraseologismen. Eine zweite Gruppe stellen die nicht- oder nur schwachidiomatischen usuellen Mehrwortverbindungen dar.

Phraseologismen definieren wir als usuelle Mehrwortverbindungen, denen die Haupteigenschaft der Idiomatizität zukommt. Sie weisen darüber hinaus folgende weitere Merkmale auf: semantisch-syntaktische Stabilität, Lexikalisierung und Reproduzierbarkeit (vgl. Fleischer 1997, S. 30). Strukturell bestehen Phraseologismen aus 1-n (auto-semantischen) Basiselementen und (fakultativ) entsprechenden Funktionswörtern. Wir plädieren für einen **engen** Phraseologiebegriff, der nun tatsächlich an die Idiomatizität im definierten Sinne, also an die Nichtübereinstimmung zwischen wendungsinternen (Teil)-Bedeutungen und wendungsexterner Bedeutung, gebunden ist. Ob eine Wortverbindung darüber hinaus ein Phraseologismus im Sinne einer usuellen, also **gebräuchlichen Wortverbindung** ist, wird für uns ebenso korpusbasiert zu entscheiden sein wie im Falle der Kollokationen. Das Kriterium der Korpusbelegung oder gar der statistischen Signifikanz für Phraseologismen anzusetzen, erscheint auf den ersten Blick problematisch: Es ist davon auszugehen, dass das Sprachgedächtnis ein ungleich größeres Repertoire dieser Wendungen besitzt als im aktuellen Sprachgebrauch tatsächlich relevant wird. Weiterhin erscheint es durchaus logisch anzunehmen, dass sich Phraseologismen auf Grund ihrer sprachstrukturellen Besonderheiten, Auffälligkeiten und Abweichungen einer statistischen Analyse wohl eher entziehen. Unsere ersten Pilotuntersuchungen ergeben jedoch ein anderes Bild: Gerade bei den statistischen Kookkurrenzanalysen las-

¹⁷ Unsere Projektgruppe kann dabei auf die weltweit größte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten, die IDS-Textkorpora, zurückgreifen. Sie sind nicht nur hinsichtlich der absoluten Größe (wie schon erwähnt: zur Zeit ca. 650 Millionen Wörter) einschlägig, sondern auch hinsichtlich der Textsortenvielfalt. Zur Zeit werden 88 externe Korpora verwaltet. Die Korpus-situation wird sich perspektivisch noch günstiger darstellen, da das IDS als Zentrum für Korpus-technologie derzeit in Kooperation mit den Universitäten Tübingen und Stuttgart ein Standardkorpus (Referenzkorpus) aufbaut, das den Entwicklungsstand der deutschen Gegenwartssprache umfassend dokumentieren wird.

sen sich bereits eine Vielzahl idiomatischer Kollokationspartner nachweisen, die sich beim Betrachten der entsprechenden einbettenden sprachlichen Umgebungen als Phraseologismen erweisen, die also über den binären Status hinausgehen und eine phrasale Festigkeit besitzen. Auch die Umkehrprobe bestätigt diesen Zusammenhang: Die Basiselemente eines Phraseologismus stellen gleichzeitig statistisch signifikante Kollokationspartner dar. Dafür seien drei Beispiele angeführt, die in der Phraseologie bereits vielfach besprochen wurden:

Suchwort	Kollokationspartner
----------	---------------------

Flinte ins Korn werfen

Flinte	Korn, werfen
Korn	Flinte, werfen
werfen	Flinte, Korn

[...] wo Barthel den Most holt

Barthel	Most, holen
Most	Barthel, holen
holen	-

Streit vom Zaun(e) brechen

Streit	Zaun, brechen
Zaun	Streit, brechen
brechen	Zaun

Eine weitere mögliche Einschränkung der Praktikabilität der statistischen Validierung könnte durch den Sachverhalt gegeben sein, dass viele Redewendungen vor allem im **mündlichen** Sprachgebrauch aktualisiert werden und mündliche Kommunikation nach wie vor in den meisten Korpora unterrepräsentiert ist. Aber auch hier bestätigen unsere Korpusanalysen die in der Kommunikationslinguistik immer häufiger vertretende These, dass – durch den Einfluss der elektronischen Medien bedingt – mündliche Elemente zunehmend in die schriftliche Kommunikation, besonders des öffentlichen Sprachgebrauchs, vordringen.

Weitere usuelle, nicht- oder nur schwachidiomatische Mehrwortverbindungen. Zu dieser Gruppe gehören solche Wortverbindungen wie (kommunikative) Routineformeln (z.B. Grußformeln wie *Guten Morgen*), Zwillings- bzw. Drillingsformeln (*Haus und Hof; Wein, Weib und Gesang*), topische Formeln (z.B. Sprichwörter, geflügelte Worte), öffentliche Formeln und Sprüche (z.B. Losungen, Slogans, bekannte Sätze wie *Wer zu spät kommt ...* usw.), Funktionsverbgefüge, typische synsemantische Konstruktionen (z.B. *und zwar, und so weiter*). Auch diese Zugehörigkeit wird sich für uns in erster Instanz aus der Korpusvalidierung ergeben.

Auf den ersten Blick erscheint die Abgrenzung zwischen Kollokationen, Phraseologismen und weiteren usuellen Wortverbindungen schwierig. Die Kombination *Haus und Hof* ist eine Kollokationsrelation, in gewissem Sinne gibt es auch eine phraseologische Bedeutung und sie stellt einen Typ usueller Wortverbindungen, eine Zwillingsformel, dar. Hier wird die Schwierigkeit rein theoretischer klassifikatorischer Einordnungen nur

zu deutlich. Unser System bietet die Möglichkeit, diese Phrase unter verschiedenen Perspektiven aufzunehmen: 1. als Kollokation, also strukturelle Relation der beiden signifikant miteinander vorkommenden Nomina *Haus-Hof* (wobei hier die Konjunktion *und* keine Rolle spielt); 2. als Phraseologismus *Haus und Hof verspielen* (hier ist die Konjunktion für die Beschreibung der syntaktischen Festigkeit der Phrase unabdingbar) und 3. als Zwillingsformel.

3. Die lexikografische Aufbereitung in einem hypertextbasierten Informationssystem

Zum besseren Verständnis ist es zunächst notwendig, einige wesentliche Inhalte unseres lexikografischen Konzepts zusammenzufassen. Das Projekt „Wissen über Wörter“ verfolgt hauptsächlich zwei Ziele:

– die Entwicklung eines auf den neuesten Stand der Technik und der linguistischen Forschung ausgerichteten **lexikografischen ‚Arbeitsplatzes‘**. Die offene, zeitlich nicht begrenzte und stets aktualisier- und erweiterbare Hypertextstruktur ermöglicht eine sehr flexible und dynamische lexikografische Aufbereitung, die stets auch durch neue ‚Module‘ ergänzt werden kann. Die Lexikografen werden dabei direkt auf die zu Grunde liegende objektorientierte Datenbank zugreifen können. Die Artikelstruktur eines Lemmas ist in folgende fünf Informationsdimensionen gegliedert:

Schreibung&Aussprache
Bedeutung&Verwendung
Grammatik
Geschichtliches&Sachliches
Sprachkritisches&Normatives

Für den Lexikografen enthalten die einzelnen Dimensionen obligatorische und fakultative Angaben, wobei alle Angaben strikt lesartenbezogen gemacht werden. Innerhalb dieser Dimensionen, die formal als komplexe Objekte oder Klammern aufgefasst werden, eröffnen sich dem Lexikografen sehr verschiedene Möglichkeiten: Zwei grundsätzliche Formen stellen objektsprachliche Ausdrücke und beschreibungssprachliche Texte dar, u.a. in Form von Listen, aus denen ausgewählt werden kann. Der Lexikograf hat aber auch die Möglichkeit – ohne den Zwang der Platzökonomie – größere narrative Angaben, Kommentare und Erläuterungen zu geben, was sich besonders für die pragmatischen, diachronen und sprachkritischen bzw. -normativen Angaben von großem Vorteil erweisen wird. Von grundsätzlicher Natur ist auch die explizite Unterscheidung zwischen kompetenzbasierten und korpusbasierten Angaben, wobei korpusbasierte Angaben zunehmende Präferenz erhalten werden. Ein wesentliches Potenzial eines solchen Hypertextsystems stellt die Verlinkungsstruktur dar, die nicht nur Verweise von einem Lemma zum anderen beinhaltet, sondern ein komplexes System von Link-Typen, u.a. als Verweise zum Glossar mit den einschlägigen linguistischen Termini, zu linguistischen Publikationen, zu Korpusbelegen, zu anderen Projekten wie dem analogen IDS-Grammatikprojekt GRAMMIS.

– die breite Nutzung und Anwendung durch Fachleute und sprachinteressierte Laien. Der Nutzer kann die Informationen je nach seinem Interesse unterschiedlich tief in einer oder mehreren der Dimensionen abrufen, die für ihn relevanten Informationen auswählen und individuell zusammenstellen (vgl. dazu ausführlich Haß-Zumkehr 2000 und die Internetpräsentation des Projektes).

Das Modell der usuellen Wortverbindungen stellt dabei ein zentrales Konzept in unserem Projekt dar, und zwar in mehrfacher Hinsicht: Korpusbasierte Kookkurrenzanalysen dienen zum einen als lexikografisches Arbeitsinstrument zur Rekonstruktion von Lesarten, prototypischen und typischen semantischen Merkmalen bzw. Eigenschaften eines Lemmas und in diesem Sinne weitergehend auch als linguistische Analysebasis. Zum anderen werden usuelle Wortverbindungen selbst zu einem eigenständigen lexikografischen Beschreibungsgegenstand. In unserem Informationssystem werden sie auf verschiedene Weise aufbereitet:

- als grammatische Angaben zu jedem Einwortlemma (z.B. als Angabe zum syntaktischen Verhalten, zur Argumentstruktur und zur Valenz),
- als semantische Kookkurrenzangaben zu jedem Einwortlemma,
- als eigenständiger Lemmatyp 'Mehrwortlemma'.¹⁸

Alle Kookkurrenzangaben basieren – wie erläutert – auf dem Signifikanzkriterium und werden durch die statistische Kookkurrenzanalyse mit flexiblem Feinheitsgrad aus den IDS-Korpora aufbereitet. Bereits diese reinen Signifikanzlisten können einigen Aufschluss über die prototypischen Zuschreibungen geben. So rangieren beim Lexem *Hund* folgende Kollokationspartner auf den vorderen Plätzen: *Katze, Leine, bellen, beißen, streunende, bunter, Herrchen*. Unser Analysetool ermöglicht es dann, dieses Kookkurrenzaufkommen weiter zu spezifizieren: nach autosemantischen Kollokationspartnern¹⁹, synsemantischen Kookkurrenzpartnern (wobei die expliziten Informationen zu den synsemantischen Relationen in der Informationsdimension 'Grammatik' abgehandelt werden, wohin man von der semantischen mit einem Link navigiert wird), nach usuellen phrasalen Mustern und prototypischen Kollokationspartnern. Die autosemantischen Partner eines Lemmas können dann weiterhin nach Wortarten geordnet werden, z.B. nach allen signifikanten Verben, Adjektiven und Adverbien, aber auch nach allen Nomina zu einem Nomen.

Verbale Kollokationspartner von *Hund* sind z.B:

bellen-beißen-streunen-wedeln-wecken-hetzen-ausführen-anleinen-spazierengehen-abrichten

Nominale Kollokationspartner von *Hund* sind z.B.:

¹⁸ Im Projekt gibt es insgesamt 3 Lemmatypen: Einwortlemma, Mehrwortlemma und Wortelementlemma (z.B. Angaben zu Affixen wie ver-, -lich usw.)

¹⁹ Bei den autosemantischen Kollokationspartnern von Einwortlemmata ist die Unterscheidung zwischen 'Basis' und 'Kollokator' (Hausmann) oder 'collocate' und 'node' (Sinclair) nicht mehr nötig, da diese syntagmatischen Angaben zu jedem Lemma erfolgen sollen. Auch die Auffassung, dass im Regelfall das Nomen die Basis und das Adjektiv den Kollokator darstellt (vgl. u.a. Hausmann 1984), wird sich immer dann relativieren, wenn das jeweilige Adjektiv das zu beschreibende Lemma darstellt, zu dem dann u.a. signifikante nominale Kollokationspartner erfasst werden.

Katze-Herrchen-Leine-Gassi-Tierheim-Halter-Schwanz-Esel-
Leinenzwang-Besitzer-Ofen-Halsband-Spaziergänger-Wild-Tier-
Pudel-Maulkorb-Postbote-Kind-Hubschrauber

Adjektivische bzw. adverbiale Kollokationspartner von *Hund* sind z.B.:

hart-gefährlich-freilaufend-tot-schlafend-arm-angeleint-bissig-
geprügelt-eingeschläfert-tollwütig-herrenlos-geliebt-scharf

Gerade diese Wortartenzuordnungen stehen in einem engen Zusammenhang mit den semantischen Angaben zum jeweiligen Lemma (u.a. der Bedeutungsparaphrase der jeweiligen Lesart, den Angaben zu semantischen Klassen und Rollen sowie den paradigmatischen Angaben, u.a. zur Synonymie, Hyperonymie und Antonymie).²⁰ Die Analyse usueller, aus Korpora extrahierter Kookkurenzen stellt für uns einen entscheidenden empirischen Zugang zu sprachlichem Wissen generell dar. Man kann in diesen Listen sowohl Indikatoren für typische semantische Merkmale und Eigenschaften eines Lemmas und den entsprechenden lexematischen Relationen finden als auch Bedeutungszuschreibungen, die das klassische Wörterbuch nicht anbieten kann.

Beim Lexem *Hund* lassen sich z.B. folgende semantische paradigmatische bzw. syntagmatische Relationen anhand seiner Kollokationspartner identifizieren²¹:

Hund-Haustier (Hyponymierrelation)

Hund-Schäferhund/Pitbull (Hyperonymrelation)

Hund-schwarz (Qualitätsrelation)

Hund-bellen, beißen (Handlungsträgerrelation/Aktorrelation)

Hund-Tierheim (Lokalisationsrelation)

Hund-Katze (Kontrastrelation)

Ein besonderes Erklärungspotenzial besitzen usuelle phrasale Muster und jene Kookkurrenzpartner, die in der weiteren Umgebung eines Wortes eine größere Anziehungskraft haben als andere Lexeme. Diese Kookkurrenzpartner können als Indikatoren für eine verfestigte Umgebung einer Kollokation angesehen werden, wie bei folgenden Beispielen. (Das erste Wort ist der jeweils statistisch signifikante Kollokationspartner von *Hund*; die folgenden kommen signifikant in der Umgebung der jeweiligen Kollokation vor.)

Hund

Katze wie
bellen Karawane
Leine gefährliche
beißen Letzten
bunter ein wie
Schwanz wedelt
Gassi ihren gehen
Halten gefährlicher Führen
Ofen hervorlocken

²⁰ Empirische Testläufe auf der Basis der IDS-Korpora haben unsere Erwartungen hinsichtlich der heuristischen Relevanz der Korpusdaten für eine Explizierung semantischen Wissens weit übertroffen. Es hat sich nämlich immer deutlicher ein systematischer Zusammenhang zwischen der Festigkeit von Wortverbindungen, also fixierten Wortkombinationen, einerseits und Bedeutungszuschreibungen andererseits herauskristallisiert.

²¹ Zu semantischen Binnenrelationen vgl. vor allem Klix 1984.

gefährliche einschläfern
 harter ein
 schlafende wecken
 geprügelter ein wie herumlaufen
 gehetzt Tränengas
 bissige Kampfhunde
 scharfen bewacht

Für den Nutzer werden alle Angaben zur signifikanten Umgebung des entsprechenden Lemmas mit den eben beschriebenen Differenzierungsmöglichkeiten z.B. nach Wortarten, phrasalen Mustern, prototypischen und idiomatischen Kollokationspartnern in Form von Listen innerhalb der Angabe 'Kookkurrenz' abrufbar sein.

Wie man an den Beispielen unschwer erkennen kann, lassen sich anhand der eben aufgeführten, durch die Kookkurrenzanalyse gewonnenen phrasalen Muster darüber hinaus **usuelle idiomatische** Wendungen identifizieren. Für das Lexem *Hund* ergibt die Kollokationsanalyse u.a. folgende Kandidaten:

Katze	[wie Hund und Katze],
Ofen	[keinen Hund hinter dem Ofen (her)vorlocken],
Karawane	[die Hunde bellen, die Karawane zieht weiter],
bellen[...], beißen,	[Hunde, die bellen, beißen nicht],
begraben	[da/hier liegt/ist der Hund begraben],
gekommen	[auf den Hund gekommen sein],
wecken, schlafend	[schlafende Hunde wecken],
getroffene	[getroffene Hunde bellen]
bunter	
armer	
harter	
toter	
geprügelter	
scharfer	
gefährlicher	
dicker	

Solche idiomatischen Kookkurrenzen werden zu Kandidaten für einen eigenen Wortartikel; mit einem Link wird in solchen Fällen auf unseren zweiten Lemmatyp, das 'Mehrwortlemma', verwiesen. Verallgemeinernd könnte man auch sagen, dass usuelle Kookkurrenzen dann zu eigenständigen Lemma-Kandidaten werden, wenn der Autor weitere Angaben für nötig hält, die über die statistische Signifikanz hinausgehen und ein weiteres ‚Nachschlagebedürfnis‘ angenommen wird.

Die Extraktion geeigneter Kandidaten für diesen Lemmatyp erfolgt in zwei Schritten: Zunächst werden innerhalb der Kollokationslisten aller Lemmata jene Wortkombinationen isoliert, die über die statistische Signifikanz hinaus eine idiomatische Qualität besitzen und bei denen sich auf Grund der Analyse der sprachlichen Umgebung ergibt, dass sie auch in ein **usuelles Umfeld** eingebettet sind. Somit stellen sie nicht nur eine usuelle Verbindung zwischen zwei Wörtern, eine Kollokation, dar, sondern werden zu einem Kandidaten für ein Mehrwortlemma. Diese Menge wird den Kern unserer Mehrwortlemmaliste bilden. Ein Abgleich mit vorhandenen Wörterbüchern und idiomatischen Sammlungen kann uns dann perspektivisch Aussagen darüber bringen, welche

festen Fügungen von unserer Analyse bis dato nicht erfasst wurden und welchen Status diese dann aufweisen.

Die **Artikelstruktur der Mehrwortlemmata (MWL)** ist in vielen Teilen mit der eines Einwortlemmas identisch. Es gibt naturgemäß darüber hinaus aber Informationen und Angaben, die speziell für Mehrwortlemmata einschlägig sind. So enthält die Informationsdimension '**Grammatik**' neben allen obligatorischen grammatischen Angaben zu den einzelnen Elementen auch Angaben zur syntaktischen Binnenstruktur der MWL (u.a. syntaktische Transformationen und Anomalien), zur Negation und Komparation sowie zur MWL-Wortart.

In die Informationsdimension '**Bedeutung und Verwendung**' werden nicht nur eine oder mehrere Bedeutungsparaphrasen des Mehrwortlemmas aufgenommen, sondern es sind auch Angaben zu Mehrwortklassen, u.a. Phraseologismus, topische Formel, Routineformel, Funktionsverbgefüge, Zwillingsformel und synsemantische Konstruktionen, enthalten. Diesen Klassen können jeweils wieder bestimmte Subtypen zugeordnet werden. Beim Phraseologismus wären dies z.B. solche Typen wie Phraseolexem (*armer/bunter Hund*), komparativer Phraseologismus (*wie Hund und Katze*) oder phraseologischer Teilsatz (*auf den Hund gekommen*). Des Weiteren sind Angaben zur semantischen Binnenstruktur des Phraseologismus, wie Anzahl und Charakter der Basiselemente (Idiomatizität und Unikalität), zu phraseologischen Reihen, phraseologischen und nicht-phraseologischen Synonymen, Antonymen, Hyperonymen usw., zu typischen Kollokationspartnern und sprachlichen Umgebungen eines Phraseologismus vorgesehen. Ein wichtiger Aspekt ist für uns die strikte Trennung zwischen kompetenzbasierten Angaben zu semantischen Variationen und Restriktionen einerseits und korpusbasierten Variations- und Modifikationsangaben zu einem Phraseologismus andererseits. Diese strikte Trennung ist nicht nur ein formales Problem, sie hat vielmehr weitreichende Konsequenzen, z.B. bei der Angabe syntaktisch-semantischer Restriktionen. So kann der Lexikograf durchaus kompetenzbasiert Restriktionen für jene Modifikationsfälle vermerken, bei denen keine phraseologische Bedeutung mehr vorliegt. Zugleich muss er aber auch Korpusbelege angeben, die typische textabhängige Modifikationen abbilden.²²

Die Möglichkeiten, die ein offenes Hypertextsystem bietet, werden besonders deutlich im Bereich der pragmatischen Angaben, die u.E. gerade für Mehrwortlemmata auf Grund ihres hohen Grades von Kontextgebundenheit von entscheidender Bedeutung sind. Die **pragmatischen Angaben** in unserem Informationssystem sind nicht mehr vergleichbar mit den bereits kritisierten kargen stilistischen Etiketten tradierter Wörterbücher. Folgende pragmatische Angaben sind sowohl für Einwort- als auch für die Mehrwortlemmata vorgesehen, wobei nicht zu jedem Lemma alle Angaben möglich und sinnvoll sind:²³

²² Zur Unterscheidung zwischen usueller Variation und okkasioneller Modifikation und ihren textuellen Determiniertheiten vgl. Burger 1998, S. 27 und S. 150ff.

²³ Viele Lemmata aus dem allgemeinsprachlichen Bereich werden beispielsweise keine spezifische Zugehörigkeit zu einer Textsorte oder zu einem Register aufweisen. Die vorgesehenen Kontextangaben aus unseren Korpora sollen jedoch eine differenziertere Einordnung präferierter kommunikativer Situationen für die Verwendung eines Lemmas möglich machen, als das in der bisherigen lexikografischen Praxis auch nur annähernd geleistet werden konnte.

Angabe zur Diskursspezifität

diatopische Angabe

→ enthält Angaben zu nationalen und regionalen Varietäten;

Angaben zur Textspezifität

→ enthält Unterangaben

zum Fach- und Sachbereich;

zur Textsorte;

zu Textfunktionen (z.B. deklarativ, obligativ usw.)

Fachsprachenangabe

Angabe zur Gruppenspezifität

(nach sozialem Stand, Geschlecht, Alter usw.)

Angabe zur Situationsspezifität

→ enthält Angaben

zur spezifischen Medialität (wie ‚mündlich – schriftlich‘);

zur spezifischen Sprecher-Hörer-Konstellation (wie ‚monologisch – dialogisch‘, ‚symmetrisch – asymmetrisch‘, ‚formell – informell‘)

zur Registergebundenheit (z.B. ‚privat‘, ‚bei der Begrüßung‘, ‚vor Gericht‘),

Angaben zu Sprechereinstellungen und -bewertungen

(gegenüber dem Denotat oder Bezugsgegenstand)

Illokutionsangabe

(bei einem satzwertigen Mehrwortlemma).

Alle diese Angaben können auch narrativ in einem Kommentar gemacht werden. Wenn ein Autor beispielsweise eine nationale Varietätenzuordnung für problematisch hält, hat er die Möglichkeit, dies genauso explizit zu formulieren wie relativierende Zuordnungen von Medialität [z.B. „bisher vorwiegend in der mündlichen Kommunikation verwendet, zunehmend auch Eingang in schriftliche Texte“].

Wie die korpusbasierte Analyse zu einer Mehrwortbeschreibung führen kann, die die klassischen Wörterbücher nicht anbieten, soll am Phraseologismus *Da liegt der Hund begraben* kurz illustriert werden. Zunächst kann man eine linguistische Einordnung ohne Korpusbasierung vornehmen. Die Bedeutungsparaphrase dieser Wendung stimmt in allen einschlägigen Wörterbüchern im Wesentlichen überein:

Allgemeinsprachige Wörterbücher

DUW: [...] R da liegt der H. begraben (ugs.; *das ist der Punkt, auf den es ankommt, die Ursache der Schwierigkeiten* [...])

PAUL: [...] *da liegt der H. begraben* (17. Jh.) >das ist der Kern der Sache< (Ursprung unklar);

WAHRIG: [...] *da liegt der ~ begraben!* *das ist die Quelle des Übels, der Kern der unangenehmen Angelegenheit; [...]*

Spezialwörterbücher

AGRICOLA: [...] *da liegt der H. begraben!* (umg.; *das ist die Quelle des Übels*). [...]

DUDEN 11: [...] **da liegt der Hund begraben** (ugs.): *das ist der entscheidende Punkt, an dem alles scheitert* [...] **wissen, wo der Hund begraben liegt** (ugs.): *den entscheidenden, schwierigen Punkt kennen, der [unüberwindliche] Schwierigkeiten bereitet: [...]*

Es handelt sich um einen satzwertigen Phraseologismus, der vollidiomatisch ist und zwei nicht-unikale, aber vollidiomatische Basiselemente besitzt (*Hund, begraben sein*). Als lexikalisch-semantische Restriktionen könnte man beispielsweise folgende Versionen angeben: **da liegt das Huhn begraben; *da liegt der Hund eingegraben*. Bei der Angabe zu semantischen Relationen seien nur die phraseologische Reihe *auf den Hund kommen, ein dicker Hund, keinen Hund hinterm Ofen vorlocken* und die phraseologischen Synonyme z.B. *[Da] liegt der Hase im Pfeffer* genannt.

Aus pragmatischer Perspektive ist das illokutive Potenzial besonders prägnant, das sich aus dem – von allen Wörterbüchern als Normalform angegebenen – satzwertigen Status ergibt. Es handelt sich um einen exklamativen Sprechakt. Diese Äußerung hat einen emphatischen Charakter, d.h., der Sprecher verleiht ihr einen besonderen Nachdruck. Die Recherche im Korpus ergibt dann ein sehr viel differenzierteres Bild. Hier seien nur vier Aspekte genannt:

Lesarten. Es gibt eine weitere Lesart dieses Phraseologismus, die in keinem Wörterbuch zu finden, aber im Korpus frequent belegt ist: ~ = ein Ort, wo absolut nichts los ist; ein ‚Nest‘ am Ende der Welt.

... *Schlaglöcher. In Nordhausen liegt der Hund wirklich begraben ...*

... *der Bär. In Huchting ist vielmehr der Hund begraben ...*

... *Auf dem Lande liegt der Hund begraben ...*

Normalform. Die Normalform *Da liegt der Hund begraben* ist schon auf Grund der Frequenz allein nicht zu rechtfertigen. Ebenso häufig ist die Form *hier liegt der Hund begraben* belegt, aber auch solche Formen wie *im Verwaltungshaushalt liegt der Hund begraben; in der eisigen Betrachtungsweise liegt der Hund begraben*

Korrekterweise müsste die Normalform so aussehen:

Der Hund liegt/ist begraben + [lokal/deiktElem]

Illokution. Damit verbunden erscheint das deiktische Element *da* fakultativ für den satzwertigen Status, was wiederum Konsequenzen für die illokutiven Merkmale hat. Der Exklamativstatus ist zumindest nicht in der Absolutheit zu halten. Da auch die häufig vorkommende Markierung durch das Ausrufezeichen völlig fakultativ ist, fungieren die meisten Belege als reine Feststellungssätze oder als Teilsätze in einem übergeordneten Satzzusammenhang.

Modifikationen. Die Korpusanalyse erbringt auch eine Vielzahl textgebundener Modifikationen, die rein intuitiv sicherlich nicht zu konstruieren wären:

Und da liegt oft der Hund begraben

Und genau da/dort/hier liegt [wohl] der Hund begraben

Hier liegt der Hund von Maastricht begraben

Hier liegt doch der Hund begraben!

Das soeben beschriebene Modell und die damit verbundenen lexikografischen Möglichkeiten entbinden uns von vielen Schwierigkeiten, mit denen klassifikatorische Ansätze unweigerlich konfrontiert werden, wenn sie sich realen Sprachbefunden zuwenden: Oft genug passen Belege aus Korpora auf Grund ihrer Komplexität – im Vergleich zu konstruierten Beispielen – nicht eindeutig oder gar nicht in die im Vorhinein festgelegten linguistischen Kategorisierungen.

4. Linguistischer Erkenntnisgewinn – ein Ausblick

Neben den empirischen Ergebnissen, die durch die Erfassung und Aufbereitung usueller Wortverbindungen des Deutschen sowohl des tatsächlichen Kollokationsvorkommens als auch des phraseologischen Bestandes der deutschen Gegenwartssprache zu erwarten sind, beziehen sich die innovativen Forschungsansätze im Bereich der Idiomatik vor allem auf folgende drei Aspekte:²⁴

- die Entwicklung eines differenzierten Kookkurrenzmodells, das sowohl eine Klassifizierung von Kollokationen als auch eine Systematisierung der semantischen Relationen und der Kontextmuster beinhaltet;
- neue Erkenntnisse zu kontextabhängigen Variationen und Einbettungen von Phraseologismen auf der Basis empirischer, korpusbasierter Analysemethoden;
- die Erfassung und Systematisierung weiterer Typen usueller Wortverbindungen, die keine Kollokations- oder Phraseologiephänomene sind und bisher wenig oder gar nicht im Fokus des Forschungsinteresses standen, da sie sich trotz ihrer Usualität der Sprachintuition des Lexikografen immer noch weitgehend verschließen.

Perspektivisch sollen sowohl die usuellen Wortverbindungen des Deutschen umfassend aufbereitet und mit linguistischen Angaben versehen als auch die entsprechenden Verwendungsmuster in systematischer Form angeboten werden, wobei dies nicht nur für Deutschlerner und Fremdsprachennutzer sondern auch für Muttersprachler und vor allem für an Idiomatik interessierte Fachkollegen von Nutzen sein wird. Im Mittelpunkt steht dabei die Frage, auf welche Art und Weise und unter welchen Bedingungen zunächst rein okkasionelle Wortverbindungen einen zunehmend usuellen Status erlangen und welchen Anteil die entsprechenden Kontexte daran haben. Für **Phraseologismen** kann dies beispielsweise dann gelten, wenn die ursprünglich okkasionell bedingten Modifikationen eines Phraseologismus durch zunehmende Frequenz und Gebrauchshäufigkeit selbst einen usuellen Charakter annehmen. Die ausgearbeitete Angabestruktur wird es auch ermöglichen, nach unterschiedlichsten linguistischen Problemen zu recherchieren, z.B. nach Herkunfts- und Bezugsbereichen usueller Wortverbindungen, durch deren Systematisierung man zu universalen kognitiven Mustern der Idiomatik im Sinne Dobrovol'skijs gelangen kann: z.B. nach allen usuellen Wortverbindungen, die in irgendeiner Weise auf 'Hauswesen' referieren, oder nach allen Idiomen, die das Erreichen eines Ziels ausdrücken (vgl. dazu Dobrovol'skij 1995).

Wie fruchtbar Fragestellungen im Bereich der usuellen Wortverbindungen für die allgemeine Sprachtheorie im Spannungsfeld zwischen *langue* und *parole* gemacht werden können, zeigt sich beim Teilbereich der **Kollokationen**. Greifen wir in diesem Zusammenhang ein Beispiel Feilkes (1996, S. 164f.) auf: *Zähne putzen* vs. **Zähne waschen* bzw. **Zähne reinigen*. Ist die Verbindung *Zähne putzen* nun eine systembedingte restringierte Fügung oder liegt die Festigkeit auf der Ebene der *parole*? Ist es 'nur' eine Kollokation oder nicht doch ein Phraseologismus? Gilt die semantische (?) Restriktion

²⁴ Zu den Möglichkeiten und Perspektiven, die das Projekt für die Erprobung und Entwicklung neuer Technologien und offener Hypertextsysteme und – damit verbunden – für die Metalexikografie eröffnet, vgl. Haß-Zumkehr 2000. Der durch die konsequent korpusbasierte Einbeziehung und Aufarbeitung der sprachlichen Kontexte der Lemmata mögliche linguistisch-theoretische und empirische Gewinn für die Pragmatikforschung, Textlinguistik und die moderne Stilistik muss an anderer Stelle gesondert dargestellt werden.

wirklich für beide Varianten *waschen /*reinigen? Wir stimmen hier mit Feilkes Problematisierung völlig überein, dass es – rein systembedingt – keine semantische Ausschlussbedingung für die Kombination *Zähne waschen* gibt.²⁵ Aber der Muttersprachler urteilt intuitiv: „Das sagt man so nicht im Deutschen“. Alle Versuche, eine semantische Komponente zu finden, die für die Selektionsbeschränkung verantwortlich scheint, sind letztendlich nicht haltbar. So könnte man annehmen, dass das Verb *putzen* die Instrumentrelation *putzen*–*Bürste* [etw. mit der Bürste reinigen] impliziert. Das übliche *Fenster putzen* widerlegt diese Annahme ebenso wie die Vermutung Feilkes, dass die Restriktion in der Referenz auf ‚reinigen von innen‘ (*putzen*) vs. ‚reinigen von außen‘ (*waschen*) liegen könnte. Der Referenzbereich des Verbs ‚*putzen*‘ ist darüber hinaus noch stark abhängig vom Dialektgebiet. Während man in Süddeutschland fast alles ‚putzt‘ („Ich muss heute noch putzen“), würde man in Berlin und weiter nördlich höchstens ‚saubermachen‘. Feilke selbst sieht seine Interpretationsversuche nicht als Regelkonstruktionen an, sondern als „Post-hoc-Differenzierung“ (ebd., S. 165). Folgerichtig muss es sich bei *Zähne putzen* um eine *usuelle* Restriktion handeln, die sich zunächst einem Regelsystem zu verschließen scheint und daher stets auch anfällig für okkasionele Modifikationen ist. Betrachten wir z.B. die Kombination *Zahn reinigen*. Die Entscheidung, dass man zu seinem Kind nicht sagt: „Reinige dir bitte deine Zähne!“ ist rein kontextgebunden. In der Alltagssituation der Familie würden Sprecher eine solche Formulierung nicht verwenden, in der eher institutionellen Situation einer Zahnarztpraxis ist sie – quasi als Fachterminus – durchaus üblich. Der Zahnarzt informiert den Patienten, dass er zunächst die *Zähne reinigt* (oder in der nominalisierten Form *eine Zahnreinigung vornimmt*). Angesichts dieser Relativität der Ableitbarkeit muss man sich die Frage stellen, ob es tatsächlich sinnvoll ist, semantische Vereinbarkeiten aus ihrer Systemhaftigkeit zu konstruieren und somit Ausschlussbedingungen für das Kookkurrenzverhalten von Wörtern zu formulieren. Kommt man nicht auch zu diesen Ergebnissen, wenn man die Perspektive umkehrt, indem man zunächst empirische Befunde erhebt, um herauszufinden, was usuell sanktioniert ist? Wenn man diese usuellen Normen dann differenziert beschreibt, muss man sich nicht für die eine *oder* die andere Zuordnung entscheiden. Hat man einen ausreichend großen Befund, was bedeutet, dass alle weiteren Belege einem der Fälle zugeordnet werden können, muss es dann doch möglich sein, zu Erkenntnissen über Regularitäten der Kookkurrenz ganz bestimmter Wörter zu gelangen. Dieses einfache Beispiel berührt einen wesentlichen Befund, der durch das Projekt verifiziert werden soll: Usuelle Wortverbindungen im Allgemeinen und Kollokationen im Besonderen lassen Rückschlüsse auf kognitive Strukturen des mentalen Lexikons generell zu, wie sie auch Feilke zieht:

„Je stärker ausgeprägt die syntagmatische Typik des Ausdrucks d.h seine Differenzqualität und strukturelle Prägnanz hinsichtlich der Ausdrucksfunktion ist, je unverwechselbarer er also ist, [...], desto zuverlässiger kann er ein bestimmtes Schema und bestimmte Funktionen indizieren.“ (1998, S. 78)

Usuelle Wortverbindungen, vor allem Kollokationen und verfestigte Syntagmen, stellen für uns in diesem Sinne Indikatoren für die Konstituierung von Lesarten dar. Mit Hilfe korpusbasierter Kookkurrenzextraktionen und -analysen wird es möglich sein, zum einen durch Empirie und nicht mehr nur durch Introspektion zu einer Systematisierung sprachlichen Wissens zu gelangen, zum anderen aber auch ‘pragmatische’ Bedeutungen

²⁵ In diesem Sinne auch Burger (1998, S. 50).

zu rekonstruieren, die das Wörterbuch in keinem Fall aufbereiten kann. Die Einbeziehung typischer Kontexte und phrasaler Muster ermöglicht neben der Beschreibung von Wortschatzwissen vor allem auch die Rekonstruktion diskursabhängiger Bedeutungszuschreibungen, diskursiver Stereotypisierungen und Konnotationen, die mit einem Wort verbunden sind.

Es sollte deutlich geworden sein, dass die einzelnen Fragestellungen innerhalb des Projektes „Wissen über Wörter“ nicht isoliert voneinander zu betrachten sind und in enger Verbindung mit generellen Forschungsproblemen der Linguistik stehen. Der linguistische Mehrwert, der dadurch erbracht werden kann, ist demzufolge so vielschichtig wie das Projektkonzept selbst. Der erhobene Anspruch, den deutschen Wortschatz in umfassender Weise zu repräsentieren, macht eine richtungsübergreifende Sicht auf das ganze System ‘Sprache’ geradezu zwingend erforderlich und öffnet darüber hinaus auch neue Perspektiven in sprachtheoretischer Hinsicht.

Literatur

- Agricola, Erhard (1982): Ein Modellwörterbuch lexikalisch-semantischer Strukturen. In: Agricola, E./Schildt, J./Viehweiger, D. (Hg.): Wortschatzforschung heute. Aktuelle Probleme der Lexikologie und Lexikographie. Leipzig. S. 9-22.
- Bahns, Jens (1996): Kollokationen als lexikographisches Problem. Eine Analyse allgemeiner und spezieller Lernerwörterbücher des Englischen. (= Lexicographica; 74). Tübingen.
- Burger, Harald/Buhofer, Annelies/Sialm, Ambros (1982): Handbuch der Phraseologie. Berlin/New York.
- Burger, Harald (1998): Phraseologie. Eine Einführung am Beispiel des Deutschen. (= Grundlagen der Germanistik; 36). Berlin.
- Dobrovolskij, Dmitrij (1995): Kognitive Aspekte der Idiom-Semantik. Studien zum Thesaurus deutscher Idiome. (= Eurogermanistik; 8). Tübingen.
- Donalies, Elke (1994): Idiom, Phraseologismus oder Phrasem? Zum Oberbegriff eines Bereichs der Linguistik. In: Zeitschrift für Germanistische Linguistik 22, S. 334-349.
- Feilke, Helmuth (1996): Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik. Frankfurt a. M.
- Feilke, Helmuth (1998): Idiomatic Prägung. In: Barz, I./Öhlschläger, G. (Hg.): Zwischen Grammatik und Lexikon. (= Linguistische Arbeiten; 390). Tübingen, S. 69-80.
- Fleischer, Wolfgang (1997): Phraseologie der deutschen Gegenwartssprache. 2., durchgesehene u. ergänzte Aufl. Tübingen.
- Fraas, Claudia/Haß-Zumkehr, Ulrike (1998): Vom Wörterbuch zum lexikalischen Informationssystem LEXXIS – ein neues Projekt des Instituts für deutsche Sprache. In: Deutsche Sprache 26, S. 289-303.
- Gréciano, Gertrud/Rothkegel, Anneli (Hg.) (1997): Phraseme in Kontext und Kontrast. (= Studien zur Phraseologie und Parömiologie; 13). Bochum.
- Haß-Zumkehr, Ulrike (2000): Zur Mikrostruktur im Hypertext-Wörterbuch. In: Lernberg, I./Schröder, B./Storrer, A. (Hg.): Chancen und Perspektiven computergestützter Lexikographie. (erscheint).
- Hausmann, Franz Josef (1984): Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen. In: Praxis des neusprachlichen Unterrichts, Jg. 31, S. 395-406.
- Hausmann, Franz Josef (1985): Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: Bergenholtz, H./Mugdan, J. (Hg.): Lexikographie und Grammatik.

- Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984. (= *Lexicographica* 3). Tübingen, S. 118-129.
- Hausmann, Franz Josef (1989): *Le dictionnaire de collocations*. In: Hausmann, F.J./Reichmann, O./Wiegand, H.E./Zgusta, L. (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Erster Teilband. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 5.1). Berlin. S. 1010-1019.
- Hausmann, Franz Josef (1993): Was ist und was soll ein Kontextwörterbuch? (Einführung). In: Ilgenfritz, P./Stephan-Gabinel, N./Schneider, G.: *Langenscheidts Kontextwörterbuch Französisch-Deutsch. Ein neues Wörterbuch zum Schreiben, Lernen, Formulieren*. Berlin/München/Wien/Zürich/New York. S. 5-9.
- Irsula Pena, Jesús (1994): *Substantiv-Verb-Kollokationen. Kontrastive Untersuchungen Deutsch-Spanisch*. (= *Hispano-Americana*; 3). Frankfurt a. M./Berlin/Bern/New York/Paris/Wien.
- Keil, Martina (1997): *Wort für Wort. Repräsentation und Verarbeitung verbaler Phraseologismen (Phra-seo-Lex)*. (= *Sprache und Information*; 35). Tübingen.
- Klix, Friedhart (1984): Über Wissensrepräsentation im menschlichen Gedächtnis. In: Klix, F. (Hg.): *Gedächtnis – Wissen – Wissensnutzung*. Berlin. S. 9-73.
- Lehr, Andrea (1996): *Kollokationen und maschinenlesbare Korpora. Ein operationales Analysemodell zum Aufbau lexikalischer Netze*. (= *Reihe Germanistische Linguistik*; 168). Tübingen.
- Palm, Christine (1995): *Phraseologie. Eine Einführung*. Tübingen.
- Porzig, Walter (1934): Wesenhafte Bedeutungsbeziehungen. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur* 58, S. 70-97.
- Sandig, Barbara (Hg.) (1994): *EUROPHRAS 92. Tendenzen der Phraseologieforschung*. (= *Studien zur Phraseologie und Parömiologie* 1). Bochum.
- Schmidt, Hartmut (1991): Sprachgeschichte zwischen Wort und Text. Über die Notwendigkeit einer historischen Wortkombinationsforschung. In: *Sprachwissenschaft in der DDR – Oktober 1989. Vorträge einer Tagung des Zentralinstituts für Sprachwissenschaft am 31.10. und 1.11.1989*. (= *Linguistische Studien, Reihe A, Arbeitsberichte*; 209). Berlin. S. 170-186.
- Schmidt, Hartmut (1995): Wörter im Kontakt. Plädoyer für historische Kollokationsuntersuchungen. In: Gardt, A./Mattheier, K.J./Reichmann, O. (Hg.): *Sprachgeschichte des Neuhochdeutschen. Gegenstände, Methoden, Theorien*. (= *Reihe Germanistische Linguistik*; 156). Tübingen. S. 127-143.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford.
- Stein, Stephan (1995): *Formelhafte Sprache. Untersuchungen zu ihren pragmatischen und kognitiven Funktionen im gegenwärtigen Deutsch*. (= *Sprache in der Gesellschaft. Beiträge zur Sprachwissenschaft*; 22). Frankfurt a. M./Berlin/Bern/New York/Paris/Wien.
- Steyer, Kathrin (1998): Kollokationen als zentrales Übersetzungsproblem – Vorschläge für eine Kollokationsdatenbank Deutsch-Französisch/Französisch-Deutsch auf der Basis paralleler und vergleichbarer Korpora. In: Bresson, D. (Hg.): „*Lexikologie und Lexikographie Deutsch-Französisch*“. (= *Cahiers d'Études Germaniques*; 35). Aix-en-Provence. S. 95-113.
- Steyer, Kathrin/Teubert, Wolfgang (1997): *Deutsch-Französische Übersetzungsplattform. Ansätze, Methoden, empirische Möglichkeiten*. In: *Deutsche Sprache* 25, S. 343-359.
- Stubbs, Michael (1997): „Eine Sprache idiomatisch sprechen“: Computer, Korpora, Kommunikative Kompetenz und Kultur. In: Mattheier, K.J. (Hg.): *Norm und Variation*. (= *forum Angewandte Linguistik*; 32). Frankfurt a.M./Berlin/Bern/New York/Paris/Wien. S. 151-167.
- Viehweger, Dieter et al. (1977): *Probleme der semantischen Analyse*. (= *studia grammatica*; XV). Berlin.
- Viehweger, Dieter (1982): Die Darstellung semantischer Vereinbarkeitsbeziehungen zwischen lexikalischen Elementen im einsprachigen Wörterbuch des Deutschen. In: Agricola, E./Schildt, J./Viehwe-

ger, D. (Hg.): Wortschatzforschung heute. Aktuelle Probleme der Lexikologie und Lexikographie. Leipzig. S. 23-41.

Viehweger, Dieter (1989): Probleme der Beschreibung semantischer Vereinbarkeitsrelationen im allgemeinen einsprachigen Wörterbuch. In: Hausmann, F.J./Reichmann, O./Wiegand, H.E./Zgusta, L. (Hg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie. (= HSK 5.1). Berlin/New York. S. 888-893.

Wirrer, J. (Hg.) (1998): Phraseologismen in Text und Kontext. (= Bielefelder Schriften zu Linguistik und Literaturwissenschaft 11; Phrasemata I). Bielefeld.

Wotjak, Gerd (1994): Nichtidiomatische Phraseologismen: Substantiv-Verb-Kollokationen – ein Fallbeispiel. In: Sandig, Barbara (Hg.), S. 651-677.

Wörterbücher

Agricola, Erhard (Hg.) (1992): Wörter und Wendungen. Wörterbuch zum deutschen Sprachgebrauch. Unter Mitwirkung v. Herbert Görner und Ruth Küfner. Überarb. Neufassung der 14. Aufl., 1. Aufl. der Neufassung. Mannheim/Leipzig/Wien/Zürich.

Benson, Morton/Benson, Evelyn /Ilson, Robert (1997): The BBI Dictionary of English Word Combinations. Revised edition. Amsterdam/Philadelphia.

DUW = DUDEN Deutsches Universalwörterbuch (1996): 3., neu bearbeitete und erweiterte Auflage. Auf der Grundlage der neuen amtlichen Rechtschreibregeln. Bearbeitet von Günther Drosdowski und der Dudenredaktion. Mannheim/Leipzig/Wien/ Zürich.

Duden 11 = DUDEN Redewendungen und sprichwörtliche Redensarten (1998): Wörterbuch der deutschen Idiomatik. Bearb. von Günther Drosdowski u. Werner Scholze-Stubenrecht. Nach den Regeln der neuen deutschen Rechtschreibung überarb. Nachdruck der 1. Aufl. Mannheim/Leipzig/Wien/Berlin.

Schemann, Hans (1993): PONS Deutsche Idiomatik. Die deutschen Redewendungen im Kontext. Stuttgart/Dresden.

Paul, Hermann (1992): Deutsches Wörterbuch. 9., vollständig neu bearbeitete Auflage von Helmut Henne und Georg Objartel unter Mitarbeit von Heidrun Kämper-Jensen. Tübingen.

Wahrig, Gerhard (1994): WAHRIG Deutsches Wörterbuch. Neu herausgegeben v. Renate Wahrig-Burfeind. Mit einem „Lexikon der deutschen Sprachlehre“. Gütersloh.

Zimmer, Rudolf (1990): Äquivalenzen zwischen FRANZÖSISCH und DEUTSCH. Theorie – Korpus – Indizes. Ein Kontextwörterbuch. Tübingen.

Dr. Kathrin Steyer
Institut für Deutsche Sprache
Postfach 101621
D-68016 Mannheim
E-mail: steyer@ids-mannheim.de